## What is claimed is:

4			
l.	Λn	annaratiic	comprising:
1.	$\Delta$ II	annaratus	COHIDHSHIE.
- •		mpp m- m- m-	T

a first plurality of queues, including a first queue and a second queue, each of the first plurality of queues for holding a plurality of pending memory requests;

one or more instruction-processing circuits, operatively coupled to the plurality of queues, that inserts one or memory requests into at least one of the queues based on a first instruction that specifies a memory operation, and that inserts a first synchronization marker into the first queue and inserts a second synchronization marker into the second queue based on a second instruction that specifies a synchronization operation, and that inserts one or memory requests into at least one of the queues based on a third instruction that specifies a memory operation; and

a first synchronization circuit, operatively coupled to the first plurality of queues, that selectively halts processing of further memory requests from the first queue based on the first synchronization marker reaching a predetermined point in the first queue until the corresponding second synchronization marker reaches a predetermined point in the second queue.

2. The apparatus of claim 1, wherein the plurality of queues is within a processor, and wherein the first queue is used for only vector

- memory requests and synchronizations, and the second queue is used for only scalar memory requests and synchronizations.
- The apparatus of claim 2, wherein the first synchronization
  instruction is an Lsync-type instruction.
- 1 4. The apparatus of claim 2, wherein the first synchronization 2 instruction is an Lsync V,S-type instruction.
- 5. The apparatus of claim 2, wherein, for a second synchronization instruction, a corresponding synchronization marker is inserted to only the first queue.
  - 6. The apparatus of claim 5, wherein the second synchronization instruction is an Lsync-type instruction.
    - 7. The apparatus of claim 1, further comprising:

1

2

1

2

3

4

5

6

7

8

9

a second plurality of queues, including a third queue and a fourth queue, each of the second plurality of queues for holding a plurality of pending memory requests; wherein the plurality of queues is in a circuit coupled to each of a plurality of processors including a first processor and a second processor, and wherein the third queue is used for only memory requests and synchronizations from the first processor, and the second queue is used for only memory requests and synchronizations from the second processor.

8.	The apparatus	of claim 7,	further	comprising:
----	---------------	-------------	---------	-------------

a second synchronization circuit, operatively coupled to the second plurality of queues, that selectively halts further processing of memory requests from the third queue based on the first synchronization marker reaching a predetermined point in the third queue until a corresponding synchronization marker from the second processor reaches a predetermined point in the fourth queue.

## 9. The apparatus of claim 1, further comprising:

a fifth queue for holding a plurality of write data elements, wherein each write data element corresponds to a memory request in the first queue, and wherein the write data elements are loaded into the fifth queue decoupled from the loading of memory requests into the first queue.

## 10. The apparatus of claim 1, further comprising:

a data cache; and

an external cache, wherein first synchronization instruction is an Lsync V,S type instruction, preventing subsequent scalar references from accessing the data cache until all vector references have been sent to the external cache and all vector writes have caused any necessary invalidations of the data cache.

## 11. A method comprising:

providing a first plurality of queues, including a first queue and a second queue, each of the first plurality of queues for holding a plurality of pending memory requests;

inserting one or memory requests into at least one of the queues based on a first instruction that specifies a memory operation;

based on a second instruction that specifies a synchronization operation inserting a first synchronization marker into the first queue and inserting a second synchronization marker into the second queue;

inserting one or memory requests into at least one of the queues based on a third instruction that specifies a memory operation;

processing memory requests from the first queue; and selectively halting further processing of memory requests from the first queue based on the first synchronization marker reaching a predetermined point in the first queue until the corresponding second synchronization marker reaches a predetermined point in the second queue.

- 12. The method of claim 11, wherein the plurality of queues is within a first processor, and wherein the inserting to first queue is for only vector memory requests and synchronizations, and the inserting to the second queue is for only scalar memory requests and synchronizations.
- 13. The method of claim 12, wherein the first synchronization instruction is an Lsync-type instruction.

2	instruction is an Lsync V,S-type instruction.				
1	15. The method of claim 12, wherein based on a second				
2	synchronization instruction that specifies a synchronization operation,				
3	inserting a corresponding synchronization marker to only the first				
4	queue.				
1	16. The method of claim 15, wherein the second synchronization				
2	instruction is an Lsync-type instruction.				
1	17. The method of claim 11, further comprising:				
2	providing a plurality of processors including a first processor and				
3	a second processor;				
4	providing a second plurality of queues, including a third queue				
5	and a fourth queue, each of the second plurality of queues for holding a				
6	plurality of pending memory requests;				
7	inserting to the third queue only memory requests and				
8	synchronizations from the first processor; and				
9	inserting to the second queue only memory requests and				
10	synchronizations from the second processor.				
1	18. The method of claim 17, further comprising:				
2	selectively halting further processing of memory requests from				
3	the third queue based on the first synchronization marker reaching a				
4	predetermined point in the third queue until a corresponding				

The method of claim 12, wherein the first synchronization

14.

1

1	synchronization marker from the second processor reaches a			
2	predetermined point in the fourth queue.			
1	19. The method of claim 11, further comprising:			
2	queueing a plurality of write data elements to a fifth queue,			
3	wherein each write data element corresponds to a memory request in the			
4	first queue, and wherein the write data elements are inserted into the			
5	fifth queue decoupled from the inserting of memory requests into the			
6	first queue.			
1	20. The method of claim 11, further comprising:			
2	providing a data cache and an external cache, wherein first			
3	synchronization instruction is an Lsync V,S type instruction; and			
4	preventing subsequent scalar references from accessing the data			
5	cache until all vector references have been sent to the external cache and			
6	all vector writes have caused any necessary invalidations of the data			
7	cache based on the first synchronization instruction.			
1	21. An apparatus comprising:			
2	a first plurality of queues, including a first queue and a second			
3	queue, each of the first plurality of queues for holding a plurality of			
4	pending memory requests;			
5	means for inserting one or memory requests into at least one of			
6	the queues based on a first instruction that specifies a memory			
7	operation:			

means for, based on a second instruction that specifies a synchronization operation inserting a first synchronization marker into the first queue and inserting a second synchronization marker into the second queue;

means for inserting one or memory requests into at least one of the queues based on a third instruction that specifies a memory operation;

means for processing memory requests from the first queue; and means for selectively halting further processing of memory requests from the first queue based on the first synchronization marker reaching a predetermined point in the first queue until the corresponding second synchronization marker reaches a predetermined point in the second queue.

- 22. The apparatus of claim 21, wherein the plurality of queues is within a first processor, and wherein the means for inserting to first queue operates for only vector memory requests and synchronizations, and the means for inserting to the second queue operates for only scalar memory requests and synchronizations.
- 1 23. The apparatus of claim 22, wherein the first synchronization 2 instruction is an Lsync-type instruction.